Measurement and Explanation of the Intensity of Co-publication in Scientific Research: An Analysis at the Laboratory Level*

Jacques Mairesse[a] and Laure Turner[b]

January 2002

Abstract – In this paper we study networks of academic researchers on an aggregated laboratory scale. We propose a measurement of the intensity of cooperation between laboratories, and attempt to account for its intra- and inter-town variations in relation to a number of characteristics: geographic distance between laboratories, specialization of laboratories, size of their scientific community, productivity, quality of their publications and international openness. Cooperative relations are identified on the basis of data on co-publication. These data concern French physicists from the Centre National de la Recherche Scientifique (CNRS), between 1992 and 1997.

*** Work under way; comments welcome.**

[a] CREST-INSEE, 15 Bd Gabriel Péri, 92242 Malakoff, e-mail: mairesse@ensae.fr.

[b] EUREQua, Université Paris I Panthéon-Sorbonne, Maison des Sciences Economiques, 106-112 Bd de l'Hôpital, 75013 Paris, e-mail: turner@univ-paris1.fr.

We wish to thank Michèle Crance, Serge Bauin and the members of the Unité des Indicateurs de la Politique Scientifique (UNIPS, CNRS), for their help in compiling the data base and their advice. Anne Crubelier, Claudine Hermann, Céline Allard-Prigent, Marie-Laure Allain, Dominique Foray and the participants of the Conference in honour of Paul David (Turin, 2000).

1 Introduction

Since the effectiveness of the scientific system has become essential in our knowledge-based economies, an important research field has opened. The challenge is to illuminate the role of science in economic dynamics and that of scientific institutions in the production and diffusion of knowledge and in its transformation into new competencies. The "new economics of science" therefore analyses questions as varied as the institutional configurations of scientific systems, the researcher job-market, the rules governing researchers' incentives, the allocation of public funds to research, and scientific policy. It thus contributes to reflection on the efficient organization of science (Stephan, 1996; Dasgupta & David, 1994).

The work presented in this paper is in keeping with the focus of the economics of science on knowledge-production, and is part of a broader study of the determinants of researchers' productiveness. We believe that membership of a dynamic laboratory that is central in research collaboration stimulates researchers' individual productivity and may be part of a process of cumulative advantages in which well-known scientists enhance their productivity and recognition by working in this type of laboratory. Given the substantial increase in the proportion of articles co-authored by scientists who may even belong to different institutions or countries (Gibbons et al., 1994), the unit of knowledge production seems increasingly to be the network of researchers.

In the economics of science, literature on the interactions that generate knowledge-production primarily concerns geographic externalities which promote the local emergence of new knowledge. Authors primarily study the existence of such externalities within industry or between public and industrial research[1], on the basis of patent citation data. Our work, by contrast, prefigures a study of knowledge externalities within the scientific institution, by means of co-publication data. We are interested to go

[1] Three of these studies can be mentioned here. Jaffe, Trajtenberg and Henderson (1993) account for the localization of knowledge externalities on the basis of patent citation data. The authors show that there is a very high probability of citing and cited patents belonging to the same geographic region. Jaffe & Trajtenberg (1998), also on the basis of patent citation data, confirm the localization of flows of knowledge on an international scale. Patents whose inventors live in the same country have a 30 to 80 percent greater chance of mutually citing one another than inventors in different countries. Jaffe (1989) highlights a close relationship in the USA between the number of patents and the importance of university research at national level. He interprets this as a sign of the existence of geographic externalities.

further than the observation of the spatial dimensions of research activity and analysed the determinants of the existence of externalities. Audretsch and Stephan (1996) have contributed in this direction, but still in the framework of relations between public research and industry. Based on data on the position of academic scientists in US biotechnology firms, they show that collaboration between firms and researchers in the same area is highly likely when the researchers aim to transfer knowledge towards to the firm[2], when their academic reputation is good, or when they are too young to belong to a geographically large network. Regional characteristics also seem to influence the strength of the relationship between scientists and firms. In the same spirit, we wish to identify different factors likely to play a part in the constitution and nature of networks of cooperation collaboration between academic researchers.

Many studies, often sociometric or bibliometric, have highlighted some of these determinants facilitating collaboration within academic research (see Katz, 1994, for a summary). They include, above all, researchers' reputation, popularity and visibility, demand for specific instruments needed for research, increasing specialization in science, and geographic proximity. But a rift exists within this literature, depending on the definition of the concept of a network (Shrum & Mullins, 1988). In one set of work the actors in networks are identified through the relations between them; they are distinguished by their position in a structured network (e.g. central position or not). Individual characteristics and intrinsic qualities predating the place occupied in the network, such as age, gender or skills, are not taken into account *a priori*[3]. By contrast, the second set of analyses is based on recognition of the qualities of the actors who have a different identity, status, capacities and strategies. In these studies it is these individual characteristics that determine the insertion of agents in networks and the morphology of interactions[4].

---

[2] By participating in the creation of the firm or as a member of the Scientific Advisory Board.

[3].The theory of graphs clearly illustrates this approach since individuals are represented as inter-related "points" or lines and columns of an adjacent matrix whose coefficients express the extent of the relations. For example, by adopting a definition of networks as a set of relations exceeding a certain density threshold, called a "clique", Blau (1973) shows the following in a group of 411 physicists: members of large networks are often young, work in new and innovative specialities, have a teaching post and are relatively well-known; by contrast, members of small networks are older, work in established specialities, in prestigious university departments, and are involved in administration. This seems to reflect the existence of a cycle in research careers, leading the most productive scientists towards the administrative elite.

[4] A part of the analysis by Cole and Cole (1973) on stratification in science is exemplary of this approach. They classify physicists in terms of different criteria such as age, prestige within university departments, productivity and scientific awards. They then measure the impact of these characteristics on the researchers' rank in the scientific system (in terms of reputation and visibility). This study is extended to the evaluation of discrimination of scientists on the basis of race, gender and religion.

Yet it becomes interesting to include in the same analysis structural and individual elements as constituents of networks and particularly of networks of innovation and knowledge trading. The emergence of knowledge and the innovation process is based on the interaction of multiple agents and institutions with diverse interests: scientists in public and private laboratories, firms, financiers, public authorities, etc. (Callon, 1999). Studying the structure of connections between actors by taking into account their specific characteristics should also afford insight into the mechanisms at play in scientific production.

In this article we present the first half of research carried out along these lines. Its level of analysis is the laboratory and group of laboratories at the geographic level of towns. We propose to measure the intensity of collaboration between researchers, which has the property of aggregating on this scale. This measurement enables us to assess the intensity of collaboration of each entity with partners in the same network. Observed differences of intensity are then explained by various factors: geographic distance between entities in the network, the specialization of entities, the size of their scientific community, their productivity, the quality of their publications, and their international openness. The second part of the research, that will follow this article, will study collaboration at a more detailed level, i.e. between researchers. It will enable us to consider additional aspects to those observed at laboratory level, and to reveal the role of productive and well-known "star" scientists in the elaboration and structuring of networks.

In this study we identify collaborative relations through co-publication data. These concern French physicists from the Centre National de la Recherche Scientifique (CNRS)[5] during the period 1992-97. We first analyse the individual impact of different factors on the establishment of collaborative relations and on the intensity of those relations. The approach adopted is descriptive and progressive. It consists first of establishing correlations between the considered variables. We then use an econometric model to quantify the relative weight of these factors in the determination of the degree of collaboration within networks.

Crane (1969 and 1972) and others (Crawford, 1971) highlight the importance of "star scientists" who constitute a pole around which research networks are formed.

[5] The Centre National de la Recherche Scientifique (CNRS) is a public organization for basic research, under the ministry responsible for research. With 25,000 employees (11,000 researchers and 14,000 engineers, technicians and administrative staff), a budget of 16 billion francs in 2001, and laboratories throughout the country, the CNRS covers all fields of knowledge. It relies on over 1,200 research and service units which employ the same number of lecturer-researchers as the CNRS's researchers.

Section 2 presents the scope of the study, the data and the characteristics of collaboration of the researchers studied. Section 3 defines the measurement of intensity of the collaboration. The results are presented in Section 4 and Section 5 is a conclusion.

2 Scope of the study and general characteristics of the collaboration

2.1 Scope of the study: collaboration between CNRS researchers in condensed matter physics

In this section we study networks of collaboration between 518 physicists at the CNRS. These scientists belong to the condensed matter section. They were born between 1936 and 1960 and were working at the CNRS in 1997[6]. The field of condensed matter was chosen for two reasons. First, its characteristics are suited to our study: its research is classified as pure basic science; journals with a sound reputation are clearly identifiable; the size of the field covered is clearly defined; and there is very little mobility among researchers. Second, condensed matter is a fast-growing field, honoured by the Nobel Prize for Physics awarded to Pierre-Gilles de Gennes in 1991, and which currently accounts for close to half of all French academic physics.

Condensed matter includes all states of matter, on various scales (atom, molecules, colloids, particles or cells), between liquids and solids, in which molecules are relatively close. Its study is based on a heritage of traditions, both experimental (crystallography, diffusion of neutrons and electrons, magnetic resonance imagery, microscopy, etc.) and theoretical (static physics). It is also prompted to develop more and more relations with industry around materials used in electronics, granulars, plastics, food or cosmetic gels, etc.

The group of physicists studied here represents a major part of all CNRS researchers in this discipline (654 in 1996). The CNRS and higher education institutions

[6] This dual criterion for the selection of researchers was based on two considerations: they had to be "not too young" so that we had a history of their publications (researchers born in 1960 had already been publishing for a few years in 1992, when they were 32 years old); secondly, it was necessary to have precise information as to the localization of researchers in laboratories, and in 1997 when we compiled the data base the most precise information was that of the same year.

(and, to a lesser extent, INRETS[*]) are the only public research institutions in this domain in France. In 1996 there were 738 condensed matter physicists in higher education (Barré, Crance, Sigogneau, 1999).

The fact that the researchers studied belong to the same institution, the CNRS, creates organizational proximity characterized by the sharing of common knowledge and implicit or explicit rules of organization that favour interaction and coordination (Rallet, Torre, 2000). Since they belong to the same institution and scientific community, researchers work in a context conducive to informal cooperation, that is, cooperation that does not involve prior definition of rules of coordination. The existence of this organizational proximity makes it possible to isolate the effects of geographic distance on collaboration. For example, two individuals are likely to coordinate their action if they belong to the same community, whether they are neighbours or far apart. By contrast, if they do not belong to the same organization they will not necessarily coordinate their action even if they are neighbours. In this case it is not possible to identify the contribution of geographic distance in an explanation of collaboration.

The indicator of collaboration that we use in this study is co-publication. It seems to be a reliable indicator of collaboration without being an exhaustive measurement, in so far as collaboration can have results other than publication. We compiled a data base of the publications of 518 physicists over the period 1992-1997. From this corpus of 7,789 articles drawn from the *Science Citation Index* (SCI)[7], we selected all the co-published articles. Apart from 21 researchers who published no article between 1992 and 1997, and four others who never collaborated (and published a total of only five articles during that period), collaboration seems to be the main mode of publication for the remaining 493 researchers. The number of articles co-authored during the period is 7,532 out of 7,789 published (or 97%). In fact, of the 493 researchers who collaborated, only 132 also wrote articles without co-authors (for a total of 252 articles during the period).

It is important to note that we study the networks "link by link", that is, couple of authors by couple of authors. In practice, this means that an article features in the data base as many times as it involves different couples of researchers. In so far as we wished

[*] INRETS = Institut National de Recherche sur les Transports et leur Sécurité

[7] The *Science Citation Index* (SCI) is produced by the Institute for Scientific Information (ISI). It is a US data base encompassing all scientific disciplines. Data in the base are drawn from over 3,200 of the most cited international periodicals. The quality of the data is remarkable and, in particular, the coverage of publications by CNRS units is satisfactory (UNIPS, 1999). Ninety-five percent of the articles are in English and English-language publications are totally covered by the *SCI*.

to qualify the density of collaborative relations, it seemed to us appropriate to consider that the more researchers are involved, the more weight an article will have[8].

Yet our study is situated not at the individual level but at the aggregated levels of the laboratory or group of laboratories situated in the same town. This means that we effected a change of scale, so that the networks of collaboration considered are networks of laboratories and towns. When two researchers who collaborate belong to different laboratories, we consider that the couple formed by their laboratories collaborates. Likewise, if researchers are located in different towns, we consider that the couple of towns in question collaborates. When collaboration between researchers takes place within the same laboratory, we talk of intra-laboratory collaboration and when it takes place within the same town, of intra-town collaboration.

The 518 physicists studied belonged to 82 laboratories dependent on the CNRS and situated in 36 French towns. This geographic distribution was relatively unequal since the researchers were concentrated essentially in laboratories in the Grenoble and Parisian regions (136 and 204 researchers respectively). We carried out a selection in terms of number of scientists employed in the towns and laboratories. For the study on the town level, we eliminated those which were "too small", with fewer than nine researchers. A total of 17 out of 36 towns was thus selected, and 470 out of 518 researchers. For the study on the laboratory level, we selected only those laboratories that belonged to the selected towns and were "big enough", i.e. had at least five researchers. We thus selected 34 laboratories and the number of researchers studied dropped from 470 to 427.

2.2 Two configurations of collaboration

The articles could associate researchers belonging to the same institution or not and to the same country or not. In our study, we used a simplified typology of the authors' affiliation. We distinguished between the 518 CNRS researchers selected for our study and "other" scientists, whether they were from the CNRS or French or foreign universities or other institutions. We thus identified two forms of collaboration: either it consisted of at least two CNRS researchers and possibly "other" researchers, or it consisted of at the most one CNRS researcher and "others". The reason for this distinction is a matter of practicality. CNRS researchers were the only ones for whom we

[8] Another alternative that seems to us less satisfactory for studying collaboration is breaking down the

knew the exact address of their laboratories[9]. To identify the localization of the "others" we had to be able to exploit the SCI reliably, which was not possible[10]. But without the localization of the "others", we were unable to grasp the geography of collaborative relations between a CNRS researcher and others. This is why the case of collaboration between a maximum of one CNRS researcher and others will be treated separately in the following discussion.

A total of 1,823 articles corresponds to the first form of collaboration (at least two CNRS researchers and others), and 5,709 articles to the second form (at the most one CNRS researcher and others). The 493 physicists who collaborate do so most often in these two modes. However, 38 physicists collaborate only with CNRS researchers on our list and never with others, and 69 researchers collaborate only with others and never with another CNRS researcher on our list. The 1,823 articles corresponding to the first mode of collaboration – "Group 1" – associate 424 of our researchers (= 493 – 69), while the 5,709 articles in the second mode – "Group 2" – involve 455 researchers (= 493 – 38). These figures are summarized in Diagram 1.

[diagram 1]

Table 1 shows the distribution of the number of articles of Groups 1 and 2 according to the number of their authors, depending on whether they are CNRS or other. The first five lines correspond to Group 1 and the last line to Group 2. The immediate observation is that in both cases the collaboration involves a large number of "other" researchers (only 82 articles are written by CNRS researchers only). Thus, for an average of 5.9 authors per article for Group 1 and 4.9 for Group 2, the average number of "other" researchers per article is 3.7 and 3.9 respectively. We also note that an article rarely involves more than two CNRS researchers, as the first line of Table 1 shows. Only 20% of the articles in Group 1 associate more than two CNRS researchers, which corresponds to an average of only 2.2 CNRS authors per article in Group 2.

[Table 1]

article into the number of couples, so that the article is counted only once.

[9] Owing to direct information from the Unité des Indicateurs de la Politique Scientifique (UNIPS), CNRS.

[10] In the SCI the number of authors recorded for an article is rarely equal to the number of addresses listed, and no key for correspondence between authors and addresses exists. It is, for example, possible that several authors out of all those collaborating on one article have the same address, in which case the address will appear only once on the list. But when the collaboration also involves other laboratories, we cannot know to which author to attribute the address. Another frequent example is that of multiple signatures, that is, one author who signs her/his affiliation to several laboratories, so that the number of addresses becomes greater than the number of authors, resulting again in a problem of attribution.

2.3 *Characteristics of collaboration*

This part of the study concerned mainly the 1,823 articles whose authors were at least two CNRS researchers on our list and sometimes others, i.e. Group 1. Four main reasons determine this choice.

Two have already been mentioned. The first is of a practical order. Since we do not have the location of "other" researchers, we cannot geographically situate the collaboration defined by articles in Group 2 (at the most one CNRS researcher among the authors). The second reason is analytical. By studying articles in Group 1, we study collaboration of couples of CNRS researchers. We thus control for the organizational proximity created by "common knowledge" of practices and know-how of the institution, that can lead to coordination even without geographic proximity or any other contextual factor. This enables us to isolate the impact of geographic proximity on collaboration.

But there is a third reason for the selection of articles in Group 1. The density of relations of co-publication between researchers must be high enough to allow the study of collaborative networks. From this point of view, the number of collaborative relations per couple corresponding to articles in Group 1 is greater than the number of relations corresponding to Group 2. For Group 2, the 5,709 articles involve 455 CNRS researchers and close to 10,000 others, a total of 17,500 couples with at the most one CNRS researcher. By contrast, in Group 1 the 1,823 articles have been written by 424 CNRS researchers and about 3,500 others, or by 880 couples with at least two CNRS researchers. Thus, the average number of articles per couple is only 0.33 for Group 2 as opposed to 2.1 for Group 1.

The last element explaining the restriction of the study to Group 1 is the fact that Groups 1 and 2 have certain comparable characteristics, which tends to suggest that the results of a study performed on all the data would not be very different from those obtained by studying the first group only. In order to show this, we limit ourselves to the articles situated at the intersection of the two groups, that is, the 6,753 articles published by the 386[11] CNRS researchers who collaborate in both modes. The first common characteristic is the frequency of the number of articles in relation to the number of

[11] Since we subtract from the 493 researchers who collaborate the 38 who never publish with "others" and the 69 who never publish with other CNRS researchers (386=493-38-69=424-38=455-69). See Diagram 1.

"other" co-authors, as shown in Graph 1. Thus, the probability that an article associates a number *n* of other researchers is the same in both groups. The second common characteristic is the degree of concentration of the number of articles, as shown in Graph 2. For both groups of articles the concentration curves merge, which shows that inequalities of productivity are similar in the two modes of collaboration. In particular, in both cases 30% of the articles are written by 10% of the most productive researchers.

[Graphs 1 and 2]

Yet the number of articles per researcher differs somewhat in the two groups. During the period under study, a researcher publishing in Group 1 (i.e. at least one other CNRS researcher) wrote an average of 9.9 articles[12], while a researcher publishing in the second mode of collaboration (with "others") wrote an average of 13 articles, that is, 30% more per annum. Graph 3 shows the cumulated distribution of productivity of the 386 researchers according to the two forms of collaboration. It appears that researchers publish articles in Group 1 less frequently. Thus, the number of researchers publishing fewer than six articles in Group 1 (55 researchers) is greater than 15% of the number of researchers publishing fewer than six articles in Group 2.

[Graph 3]

3 Measurement of intensity

Inclusion of agents in networks is determined by "intrinsic" individual qualities such as age, gender, skills and strategy, and by more structural variables such as number of relations they develop, geographic distance, etc. As a result, the form and functioning of networks differ. If the actors were not differentiated and if they collaborated in an equiprobable way with all the others, we would expect to observe a uniform structure of relations between all the individuals. We take this case of homogeneity as a reference. At the aggregated level of laboratories and groups of laboratories (towns) on which this work is situated, the case of homogeneity corresponds

[12] Note, this mean does not correspond to the simple mean, calculated as the ratio between the number of articles (1,823) and the number of researchers (386). The idea is to count each article as many times as there are CNRS authors, in order to attribute to each author her/his stock of publications. The sum

to the configuration in which the frequency of collaboration of each entity with all the others is the same, irrespective of their geographic localization and the characteristics of the entities.

For each entity, we compare the network that we effectively characterise on the basis of the data, to the network that would be observed in the case of perfect homogeneity of the entities. This comparison is made by means of an original measurement of intensity of collaboration between two entities, taking as a reference the situation corresponding to the homogeneous case.

3.1 Definition

In this and the following section, we assume, to simplify, that collaboration always involves at the most two CNRS researchers[13]. The network studied has a finite number of entities. It contains *N* researchers who can form *C* collaboration couples (with, by definition, C=N(N-1)/2). Let *n* be the total number of articles produced by collaboration between *N* researchers, and *p* the frequency of the number of co-publications per couple for the network, i.e. the ratio between the total number of articles, *n*, and the number of possible couples, *C*.

Consider two entities X and Y in this network. N_X researchers work in entity X and N_Y researchers in entity Y. The number of possible couples within X is C_X , and within Y is C_Y , and the number of possible couples that can be formed between researchers from X and Y is C_{XY}. Likewise, the total number of articles written jointly within X and Y is respectively n_X and n_Y , and the total number of articles written between researchers in X and Y is n_{XY}. The frequency of collaboration observed between the two entities X and Y is p_{XY} . This is the ratio between the total number of articles written by researchers in X together with researchers in Y, and the number of possible couples of researchers from the two entities. Similarly, the observed frequency of collaboration within the same entity X or Y, noted as p_X or p_Y , is the ratio between the total number of articles written together by researchers from X or Y, and the number of possible couples of researchers in entity X or Y.

of individual stocks of publications is thus 3,813. We thus have a weighted mean per number of CNRS authors per article: 3,813 / 386=9.9 article per researcher, on average.

[13] This point is discussed in Section C.

Thus, the observed frequency of collaboration within entities X and Y, and the observed frequency of collaboration between X and Y are:

$$p_X = \frac{n_X}{C_X}, \quad p_Y = \frac{n_Y}{C_Y}, \text{ and } \quad p_{XY} = \frac{n_{XY}}{C_{XY}}$$

We therefore define the intensity of collaboration as the ratio of observed frequencies to the frequency obtained for the entire network *p*. We thus obtain intra- and inter-entity intensities:

$$i_X = \frac{p_X}{p}, \quad i_Y = \frac{p_Y}{p}, \quad \text{and} \quad i_{XY} = \frac{p_{XY}}{p}$$

The entities of the network studied are homogeneous when the proportion of articles per possible couple is the same in each entity and in the network. This results from the fact that relations between individuals are aggregated at the level of the entity, and that these relations are identical and uniformly link individuals when they are undifferentiated and situated at the same geographic distance. The frequency for the entire network, *p*, is then interpreted as an expected frequency of collaboration: it is the frequency that we expect to find for X and Y if they are homogeneous. En the case of homogeneity of entities, the following equality can be verified:

$$p_X = p_Y = p_{XY} = p \text{ or } \qquad i_X = i_Y = i_{XY} = 1$$

Thus, if the entities of the studied network are homogeneous, the intra and inter intensities of collaboration are all equal to 1. Otherwise – and this is the case with our data – factors that have to be identified modify the form of the network.

The structure of intensity of a network of E entities can be represented by means of a symmetrical matrix ExE with coefficients that are either positive or nil, and of which the diagonal terms are *intraentity* intensities, and the lines (or columns) *inter-entity* intensities[14].

[14] Note that this matrix is close to the adjacency matrix used in the graph theory. The coefficients of the adjacency matrix are equal to 1 when there is a link between the entities represented on the lines and those represented in the columns; otherwise it is 0. The adjacency matrix is then a representation of indicators of inter- or inter-entity collaboration, that allows one to take into account neither the number of authors per article nor the strength of the ties. By contrast, the matrix of intensities provides qualitative information on the collaboration.

3.2 Properties of aggregation

Intensity as defined above has the advantage of being easy to aggregate. In order to see this, take X and Y, the two laboratories in town V. The total number of articles written jointly in V is the sum of articles written by laboratory X, by laboratory Y, and by the couple of laboratories X and Y. Likewise, the number of theoretical couples of researchers in V is the sum of the theoretical couples of researchers in laboratory X, laboratory Y, and the two together. We have:

$$\frac{n_V}{C_V} = \frac{n_X + n_Y + n_{XY}}{C_X + C_Y + C_{XY}}$$

This formula can also be written as follows:

$$\frac{n_V}{C_V} = \frac{n_X}{C_X} \text{ x } \frac{C_X}{C_V} + \frac{n_Y}{C_Y} \text{ x } \frac{C_Y}{C_V} + \frac{n_{XY}}{C_{XY}} \text{ x } \frac{C_{XY}}{C_V}$$

or

$$p = p_X\left(\frac{C_X}{C_V}\right) + p_Y\left(\frac{C_Y}{C_V}\right) + p_{XY}\left(\frac{C_{XY}}{C_V}\right)$$

We thus have a formula of aggregation, generalized to a network of E laboratories. By dividing the formula by p, we obtain the intensity of collaboration of a network as the weighted sum of intra- and inter-entity intensities. The weightings represent the weight of each laboratory in the network in terms of possible couples of collaboration:

$$\sum_{I} i_I w_I + \sum_{I,J\neq I} i_{IJ} w_{IJ} = 1 \qquad \text{où} \quad w_i = \frac{C_i}{C} \qquad \text{et } w_{IJ} = \frac{C_{IJ}}{C} \qquad \text{avec} \sum_{I,J} w_{IJ} = 1$$

3.3 Problems of weighting

Until now, we have supposed that the articles were written by two researchers at the most. In reality, they were also written by threesomes, foursomes, etc. By counting the number of articles written by these groups we would have obtained the total number of co-authored articles in the network. But, as indicated, we consider only couples of collaboration, and the data base repeats an article as many times as there are couples of different authors. In other words, the total number of co-authored articles is a number weighted by the number of couples that contribute to its publication.

For example, for an article published by three CNRS researchers, one belonging to an entity X and the two others to an entity Y, we would have counted only one collaborative link between X and Y, which amounts to counting the article only once, if we had reasoned in terms of threesomes. Instead, when we consider only couples of collaboration, we count three links – two between X and Y and one within Y – which amounts to repeating the article three times, as many times as the number of couples that actually contributed[15].

For the aggregation formula to remain valid in the case where more than two researchers co-author an article, it is necessary to operate this weighting of the total number of articles by the number of couples involved.

3.4 Practical calculation: an example

We have chosen to retain only those couples that have enough collaborative relations for the study of their intensity to be relevant. In the study on the town scale we therefore exclude those couples of towns linked by fewer than five articles. Likewise, in the study on the laboratory scale, we exclude couples of laboratories that collaborate on fewer than five articles. This does not reduce the number of towns and laboratories considered; it simply means that collaborations of couples are taken to be zero when they have under five articles.

Let us take the concrete example of Marseilles to describe the calculation of the intensity of collaboration, with reference to Table 2 which also presents the results of this calculation for the other towns studied. Marseilles is a town with 18 physicists on our list (column 1). There are 34 collaborations within the town (column 3) and 18 with other towns (column 4), of which ten are with Grenoble and eight with Strasbourg. Marseilles also has relations with Poitiers, Gif-sur-Yvette, Orsay, Toulouse and Villeurbanne, but these are not taken into account because they are all under six. The number of possible couples of researchers working in Marseilles is 18*17/2, or 153. The observed frequency of the number of collaborations per couple in Marseilles is therefore 34/153 or 0.2222. Given that the number of researchers in Grenoble and Strasbourg is 105 and 14,

[15] We could have considered counting one third of the link three times, that is, breaking up the article into the number of couples, so as to count the article only once. But in so far as we wished to highlight the density of collaborative relations, it seemed appropriate to consider that the more authors it has, the greater the weight of an article.

respectively, the number of possible couples of researchers linking Marseilles and Grenoble is 1,890 (105*9), and linking Marseilles and Strasbourg is 252 (14*18).

The observed frequency of the number of collaborations per couple between Marseilles and Grenoble is therefore 0.0053 (10/1,890) and between Marseilles and Strasbourg 0.0317 (8/252). In order to obtain the intensities of collaboration, we have to calculate the frequency of collaboration per couple for the set of 17 towns. This consists of the ratio between the total number of weighted articles, 2,480, and the number of possible couples that can be formed by the 470 researchers in that set, i.e. 110,215 couples (470*469/2). We thus have p=0.0225. In case of homogeneity, the frequency is that which would have been obtained for intra- and inter-Marseilles collaboration. In reality, the observed intra-frequency is much higher (0.2222) and the frequency of collaboration between Marseilles and Grenoble is lower (0.0053). The frequency of collaboration between Marseilles and Strasbourg is closer to the reference (0.0317).

Intra- and inter-entity intensity are the ratio between the observed frequencies of collaboration and p. Intra-town intensity for Marseilles is therefore 9.88 (column 5). Intensity between Marseilles and Grenoble is 0.2351, and between Marseilles and Strasbourg 1.4109, that is, a (simple) mean intensity of collaboration between Marseilles and its partners of 0.82 (column 6). In column 7 we have the (simple) mean intensity of collaboration between Marseilles and all the other 16 towns.

[Table 2]

4 Results

4.1 Intensity of collaboration in networks of French towns

Consider Table 2 again. The second column gives the number of partners in the collaboration, for each of the 17 towns. This number is 4, on average, which is low. Three towns, Poitiers, Orleans and Talence, develop no significant relations with the others.

Intensity of intra-town collaboration is presented in column 5 of Table 2. It is high and is always greater than 1, which means that the number of intra-town collaborations is much higher than the number of collaborations expected in case of homogeneity. The importance of collaboration within the perimeter of the town therefore appears clearly.

The intensity of inter-town collaboration is presented in columns 6 and 7 of Table 2. Of the 136 possible couples of towns, 34 do effectively collaborate. Column 6 indicates the mean intensity of collaboration of each town with its effective partners, and column 7 presents the mean intensity of collaboration of each town with the 16 other towns. The mean intensity of collaboration of each town with its partners is low since it is 1.01, on average, and the intensity of collaboration with all the other 16 towns is almost always lower than 1 and is 0.25 on average. On the whole, towns collaborate less with one another than what one would expect in the case of homogeneity. Moreover, inter collaboration is far less intense than intra collaboration.

The last observation concerns the wide dispersion of the mean intensity, in the case of both inter and intra. The following section explains these differences of inter-town intensity.

4.2 Determinants of the intensity of collaboration

This section presents a statistic approach to the determination of factors influencing the intensity of relations in networks of collaboration. We first implemented the approach on the level of towns and then on that of laboratories. Six factors are studied: geographic distance, specialization, size of the scientific community, productivity, quality of publications, and openness towards foreign countries (Table 3)[16].

[Table 3]

Table 4 gives a general idea of the results obtained on the town scale. It accounts for each factor studied and its quantitative impact – i.e. on the establishment of collaboration between towns – and qualitative impact – i.e. on intensity. Three correlations are thus presented. The first column presents the correlation of different factors with intra intensity, the second the correlation of these factors with a variable indicating the existence of collaboration between towns, and the third the correlation with the intensity of collaboration between towns that do actually collaborate. It is tricky to define a characteristic of a couple. Most often we use a half-characteristic, that is, the half sum of its values for each partner in the couple, or the minimum or maximum value of the characteristic.

[Table 4]

Figures 1 to 6 complete Table 4 by representing the intensity of collaboration between towns and partners in relation to the different characteristics of the couples studied. Each time we represented the most significant relation between the intensity and the characteristic the of couples, and this characteristic is then, depending on the figure, the half-characteristic, the minimum or maximum.

[Figures 1 to 6]

The results obtained on a laboratory scale are presented in Tables 5a and 5b. We have distinguished the results according to whether the laboratories belong to the same town or not. Table 5a presents the correlation of these factors with the variable indicating the existence of inter-laboratory collaboration, depending on whether the laboratories are part of the town or not. The first column in Table 5b indicates the correlation of the different factors with the intra intensity, while the second and third describe the correlation of the factors with the intensity of collaboration between laboratories that do actually collaborate, according to whether the laboratories are in the same town or not. En general, the results for towns and laboratories are similar, with the notable exception of the correlation with geographic distance and with specialization.

[Tables 5a and 5b]

4.2.1 Distance

Towns' readiness to collaborate with distant partners varies. Table 3 indicates that the average distance of a town from its partners can vary widely (the distance may be three times more in some cases than in others). For example, Montpellier collaborates with towns over a distance of 520 km on average, while Gif-sur-Yvette has relations with towns close by, situated at an average distance of 170 km. Four of the five towns situated less than 300 km from their partners are in the Parisian region (consisting of six towns in which our physicists are present).

[16] Specialization does not appear in Table 3 because its measurement involves pairs of towns. The reader is referred to Section 4.2.2.

Geographic distance plays no part in the establishment of collaboration between towns, and has no impact on the intensity of inter-town collaboration. Thus, it is not because towns are far apart that they are less likely to collaborate. Coefficients of correlation are not high, as shown in Table 4.

Yet when we scale down the study to the laboratory level, we observe a slightly negative relation between distance and the creation of collaborative relations (Table 5a, column 2). The high values of the intensity of collaboration within towns (Table 2) already suggested that immediate proximity, within the perimeter of the town, favoured collaboration. This led us to assume that immediate proximity plays a part in collaboration, and that beyond the perimeter of the town the effect of distance is slightly unfavourable to the establishment of collaborative relations. Table 2 bis illustrates this intuition by showing that the mean intensity of intra-laboratory collaboration is greater than the mean intensity of collaboration between laboratories situated in the same town, and that it itself is greater than the intensity of collaboration between laboratories in different towns.

There seem to be two types of distance: immediate proximity that favours face-to-face interaction between researchers, and distance that shrinks with the development of communication technologies. In studies on knowledge flows between public laboratories and industry, it also appears that knowledge at the origin of externalities is more easily transmitted during face-to-face interaction between the actors concerned because it is often of a largely tacit nature (Zucher, Darby and Armstrong, 1994). When relations are no longer face-to-face, distance is no longer a relevant factor in choosing collaborators. New communication technologies have certainly helped to reduce the role of geographic distance by facilitating the codification of tacit know-how and allowing its diffusion[17] (consulting data bases, reading working papers, sending articles and data, e-mail, etc.).

The impact of distance cannot, however, be translated in terms of intensity of collaboration between laboratories since, as indicated in Table 5b, the coefficient of correlation between the two variables is not significant.

---

[17] It would be interesting to know whether the recent period has favoured inter-town publication more than preceding periods. If so, a possible interpretation may be that new communication technologies have made it possible to establish new collaborative relations between distant people. We analysed our data in order to test this hypothesisand observed no clear evolution in the intensity of collaboration between researchers in distant towns over time. However, one element does qualify this result: by construction, the base allocates to each researcher her/his laboratory address in 1997 for the preceding years. The mobility of researchers is therefore not taken into account, and even if it is weak every year, it can distort the distribution of mean intensity of inter-town collaboration.

4.2.2 Specialization

We have tried to take into account the specialization of laboratories. The map of France presented in Annex 2 suggests the influence of specialization on the geography of networks of co-publication. While distance plays a very small part in the intensity of relations, it seems that this is also because collaborations are governed by this specialization of laboratories. In particular, storage rings[18], very large facilities used by physicists of condensed matter, are present essentially in two laboratories, at Orsay and Grenoble, which consequently appear to be central poles in the collaboration.

There are no "typical" specialization data traditionally used in research. We have defined a profile of specialization of entities, based on the main theoretical and/or experimental sub-domains of the discipline to which their journals belong. Classification of journals in these sub-domains is relatively well identified; it is carried out by the SCI. During the period 1992-97, the main sub-domains in which the entities under study published were physics-chemistry, general physics, solid state physics, applied physics, materials science, and crystallography. We call the vector of proportions of publications in each of these six main sub-domains, and in the other sub-domains grouped under the label "other", the "specialization profile of an entity".

The idea here is to see to what extent two entities with the same specialization profile are most likely to collaborate. For this purpose we measured the "proximity" of specialization profiles of entities two by two, and noted the correlation of this measurement with the probability of collaboration and with its intensity. To that end we used the Chi-Square statistic that allows the simultaneous comparison of several distributions of frequencies. This measurement can be interpreted in the following way: the higher the value, the more similar the profiles of specialization of the two entities considered will be. Thus, if common specialization favours collaboration between the two

---

[18] Storage rings are used to curve or oscillate the trajectory of light charged particles (electrons or positrons) that then emit "synchrotron radiation". This constitutes an extraordinary source of radiation of varying wavelengths, especially X-rays, and has become of great practical importance. Several rings have been built throughout the world for synchrotron radiation, the most recent of which have a circumference of about 500 meters. The USA has about ten rings and France has the "Super-ACO" and "DCI" rings situated at Orsay (at the LURE). The LURE has a total of 50 different experimental apparatus available for most of synchrotron radiation applications. About forty can work simultaneously. About 30 outside laboratories collaborate on a permanent basis with the LUREe, as do 20 industrial partners, in the field of physics but also chemistry, biology and environmental science, micro-production, lithography and astrophysics. The LURE rings will soon be replaced by the "SOLEIL" ring (in 2005), that will constitute a sort of "super" synchrotron radiation, that is, several thousand times brighter, and will thus afford possibilities for new applications in many scientific areas. The facility will be located at Saint-Aubin, near Orsay. The European ring of the ESRF (European Synchrotron Radiation Facility), owned by 16 countries, is situated at Grenoble and employs about 500 persons on a permanent basis.

entities, we can expect to see a positive correlation of the Chi-Square measurement with the indicator of the existence of collaborative relations and with the intensity of the collaboration.

Secondly, we wanted to distinguish those main sub-domains that favoured collaboration most. We therefore broke down the specialization profile of each entity into seven sub-profiles. Each of these sub-profiles corresponded to the degree of specialization of the entity in one of the sub-domains of the discipline or in the category "other". We again used the Chi-Square measurement to calculate the distance between each sub-profile of two entities.

Table 3 gives the average distance, in terms of specialization, between each town and its partners, measured by means of the Chi-Square statistic. To simplify the interpretation, this measurement was normalized by the theoretical value of the Chi-Square statistic[19]. It is therefore equal to 1 when the towns have the same specialization profile, and increases with the distance in terms of specialization between the towns. Bagneux and Toulouse are therefore towns that collaborate most with entities with a similar specialization profile (the normalized measurement is roughly 3), unlike Grenoble, Marseilles and Talence (where the measurement is roughly 11). On average, the (normalized) distance between partner towns as regards their specialization is 6.1.

Tables 4, 5a and 5b present the results of the correlations. The first line of the section "Specialization" corresponds to the Chi-Square calculated for the specialization profile of all the entities. The following lines include the Chi-Square calculated for each of the seven sub-profiles of the entities. On the scale of the towns, two towns specialized in physics-chemistry and in general physics are less likely to collaborate. But common specialization has no effect on the intensity of collaborative relations.

On a laboratory scale, specialization seems to favour the establishment of collaborative relations between laboratories in the same town. A laboratory with the same overall profile and a fortiori the same sub-profile as a laboratory situated in the perimeter of its town, has good likelihood of collaborating with it (Table 5a). In the case where the laboratories belong to different towns, results are less clear-cut and less meaningful. By contrast, in all the sub-domains except those of physics-chemistry and applied physics, common specialization impacts negatively on the intensity of

[19] We find it in the table of the Chi-Square law, for the degree of freedom consistent with our number of series (17) and sub-domains (7), and for a precision of 5%.

collaboration, especially in those cases where the laboratories that collaborate are not in the same town (Table 5b).

1.1.1 Size

Three indicators of size are used. The first is the number of researchers present in the towns studied. The size of the scientific community is unequal in the different towns (Table 3). Researchers are concentrated essentially in and around Grenoble and Paris (20% and 17% of all researchers, respectively). The second indicator of size, particularly well-suited to the definition of our measurement of intensity, is the number of possible couples of researchers in the towns under consideration. The last indicator of size is the stock of publications between 1992 and 1997 in the towns studied. In this respect, Grenoble is the main town because it accounts for 26.5% of the total stock of publications. Paris is second, with 14% and Orsay third, with 13% (Table 3).

Tables 4, 5a and 5b indicate the correlations of the different indicators of size with the indicator of collaboration and the intensities of collaboration. They indicate that, the greater the size of the entity (town or laboratory), the more the entity develops collaborative relations with the other entities (Tables 4 and 5a, column 2). But these relations are relatively weak, as revealed by the negative sign of correlation between the size and intensity of collaboration between partner entities (Table 5b) and Figure 1. Moreover, size does not really have a significant effect on intra-entity intensity (Table 5b, column 1).

4.2.4 Productivity

The productivity of researchers in an entity is defined as the ratio between the stock of publications[20] that appeared between 1992 and 1997 and the number of researchers in the entity. As shown in Table 3, towns are not productive to the same extent since the number of articles per researcher in the six towns ranges from 6.3 in Orleans to 36.4 in Bagneux. The mean is 15.7.

Intra-entity intensity is strongly correlated with productivity and the relationship is strong. In the case of towns, correlation of the variable indicating collaboration is strong with the least productive but not with the most productive. This result seems intuitive: a collaboration couple is formed on condition that the two towns

[20] All publications, including those with only one author or which belong to Group 2 of co-publications. Moreover, each article is counted only once, irrespective of the number of authors.

satisfy a certain level of productivity. Once this collaboration is established, the intensity increases along with the productiveness of the towns, as represented in Figure 3. In the case of laboratories, the probability of collaborative relations developing and the intensity of these relations is positively correlated to the productivity of the laboratories, without any minimum threshold constraint.

1.1.1 Quality of publications

The quality of the stock of publications by researchers in an entity is the mean of the scores given for the impact of the journals in which the articles appeared. The impact score of a journal is equivalent to the average citation rate of its articles and therefore gives information on the journal's reputation and visibility. With the selected impact measurement, citations are recorded over a period of two years. Table 3 thus shows that the towns studied publish in journals whose articles are cited an average of three times in two years. The quality of publications is variable, depending on the town. It is lowest for Poitiers, where the average citation rate of articles in journals is 2.34, and highest for Palaiseau where the average citation rate is 4.77.

The quality of publications does not impact on the intra intensity of publication. In the case of inter-town and inter-laboratory collaboration, as for productivity, the correlation of the maximum of the qualities with the variable indicating collaboration is not significant, whereas that with the minimum is. This shows that the two entities must have a stock of publications of a high quality if collaboration is to be stimulated. By contrast, the intensity of collaboration between entities seems to be independent of the quality (as shown in Figure 4 for towns). Inter-laboratory collaboration within the same town is less probable when the quality of publications by the laboratory is high.

4.2.6 International openness

We recall that it is impossible to attribute an address to those researchers called "other". It is interesting to note what this term "other" covers, and especially to see how it refers to countries other than France. We therefore built an indicator of the presence of names of foreign countries in the variable containing addresses of the laboratories of the authors of the articles studied. With a margin of error related to the non-standardization of the variable "address" in the SCI, we thus obtained information on the proportion of articles involving foreign collaboration. Columns 5 onwards in the table in Annex 1 present the number and proportion of articles written with foreign collaboration for each town. On average, towns collaborate with other countries in 30% of their articles. Paris tops the list (50%) while Bagneux has the least cooperation with foreigners (12%) (Table 3).

We interpret the proportion of articles involving a foreign country as an indicator of international openness of the laboratories and towns under study. We therefore calculate the correlation between the intensity of intra-and inter-town collaboration and this variable of openness (Tables 4 and 5). Openness and intra-intensity are not significantly correlated. By contrast, openness towards foreign countries favours the existence of collaborative relations, although it does not seem to increase the intensity of collaboration between partner entities (Figure 6).

4.3 Results of the econometric approach on a laboratory scale

A series of linear regressions have been completed in order to assess the robustness of the relations enlightened by the statistical correlations (*table 6*). The factors that appear to significantly foster collaboration are the size and productivity of the laboratories, their specialization and eventually their localization.

The probability that two laboratories collaborate increases with their size. But the impact is relatively low: if the size of each laboratory was doubled, it would result in a number of links between labs 2.8% higher in the " inter towns" case, and 10% higher in the "intra town" case. On the contrary, the intensity of collaboration diminishes with the size of the couples, and the effect is important. As a matter of fact, if the size of each lab was doubled, the collaborative intensity between labs inter towns would be 30% lower. However, the intensity of collaboration between laboratories located within the same town is not affected by the size.

In all cases, the laboratories' productivity enhances the probability that two laboratories collaborate as well as the collaborative intensity. This effect has the same order of magnitude than the size effect. If the productivity of each laboratory was doubled, it would result in a number of links between labs 2.6% higher in the " inter towns" case, and 8% higher in the "intra town" case, and the collaborative intensity between labs would be 30% higher.

The effect of the specialization profile appears to be significant in general, except for the 'intra town probability to collaborate'. The effect is negative: the closer the entities are in terms of general specialization profiles, the less they collaborate, and the lower is the intensity of their links. This result might have the following interpretation. Specialized collaborations exist among the researchers who belong to the same laboratory, whereas a certain amount of complementary favours inter laboratories and distant collaboration.

The distance has a small negative impact on the probability that two laboratories collaborate, but does not influence the intensity of collaboration. For instance, if the distance between each laboratories was doubled, the number of links inter laboratories would be 0.8% lower.

5 Conclusion

This study aimed at quantifying the impact of several factors on the shape and collaborative dynamism of scientific networks. The measure of the collaborative intensity that we proposed allowed us to identify the size of the laboratories, their productivity, their specialization and the proximity of the researchers within the same entity as the main determinants of the collaboration and of the collaborative intensity on a laboratory scale.

As was mentioned, the work will be extended by an analysis of the contribution of researchers' individual characteristics (such as age, gender, promotion, reputation) to the construction of networks. It will lead us to study the "star" scientists and their role in the elaboration of collaboration links. An econometric model should integrate these individual factors as well as the structural determinants identified in the present paper and measure their respective contributions.

6 Bibliography

AUDRETSCH D., FELDMAN M., 1996, « R&D Spillovers and the Geography of Innovation and Production », American Economic Review, 86(3), 630-640.

AUDRETSCH D., STEPHAN P., 1996, « Company-Scientist Locational Links: the case of biotechnology », American Economic Review, 86(3), 641-652.

BARRE R., CRANCE M., SIGOGNEAU A., 1999, « La Recherche Scientifique Française : situation démographique », Etudes et dossiers de l'OST, 1.

BLAU J., 1973, « Patterns of Communication among Theoretical High Energy Physicists », Sociometry, 37, 391-406.

BEAVER D., ROSEN R., 1978, «Studies in Scientific Collaboration. Part 1. The Professional Origins of Scientific Co-authorship », Scientometrics, 1, 65-84.

BEAVER D., ROSEN R., 1979, «Studies in Scientific Collaboration. Part 2. Scientific Co-authorship, Research Productivity and Visibility in The French Scientific Elite », Scientometrics, 1, 133-149.

CALLON M., 1999, « Le Réseau comme Forme Emergente et comme Modalité de Coordination : le Cas des Interactions Stratégiques entre Firmes Industrielles et Laboratoires Académiques », in Callon M. et al., *Réseau et coordination*, Economica.

COLE J. and COLE S., 1973, « Social Stratification in Science », Chicago : U. of Chicago Press.

CRANE D., 1969, « Social Structure in a Group of Scientists: a test of the invisible college hypothesis », American Sociological Review, 34, 335-352.

CRANE D., 1972, « Invisible Colleges: diffusion of knowledge in scientific communities », Chicago: U.Chicago Press.

CRAWFORD S., 1971, « Informal Communication among Scientists in Sleep Research », Journal of the American Society for Information Science, 22, 301-310.

DASGUPTA P., DAVID P., 1994, "Toward a New Economics of Science", Research Policy, 23(5), 487-521.

GIBBONS M. et al., 1994, « The New Production of Knowledge », Sage.

JAFFE A., 1989, « Real Effects of Academic Research », American Economic Review, 79(5), 957-970.

JAFFE A., TRAJTENBERG M., 1998, « International Knowledge Flows: evidence from patent citations », NBER Working Paper 6507.

JAFFE A., TRAJTENBERG M., HENDERSON R., 1993, « Geographic Localization of Knowledge Spillovers as Evidenced from Patent Citations », Quaterly Journal of Economics, 108(3), 557-598.

KATZ J.S, 1994, « Geographical Proximity and Scientific Collaboration », *Scientometrics*, vol.31, 1, p. 31-43.

LEAMER E.., STORPER M., 2000, « The Economics of Geography at the Internet Age », mimeo, University of California at Los Angeles and Management and Public Policy, November.

RALLET A., TORRE A., 2000, « Is Geographical Proximity Necessary in the Innovation Networks in the Era of Global Economy? », mimeo, Université Paris Dauphine et Institut National de la Recherche Agronomique.

STEPHAN P., 1996, "The Economics of Science", Journal of Economic Literature, 34, 1199-1235.

SHRUM W., MULLINS N., 1988, « Network Analysis in the Study of Science and Technology », *in Handbook of quantitative studies of science and technology,* ed A.F.J. Van Raan, North Holland.

UNIPS,1999, « Les Publications des Laboratoires du CNRS et leur Impact », éd. CNRS, mars.

ZUCKER L, DARBY M., ARMSTRONG J., 1994, « Intellectual Capital and the Firm : the technology of geographically localized knowledge spillovers », NBER Working Paper 4946.

[Annexes]

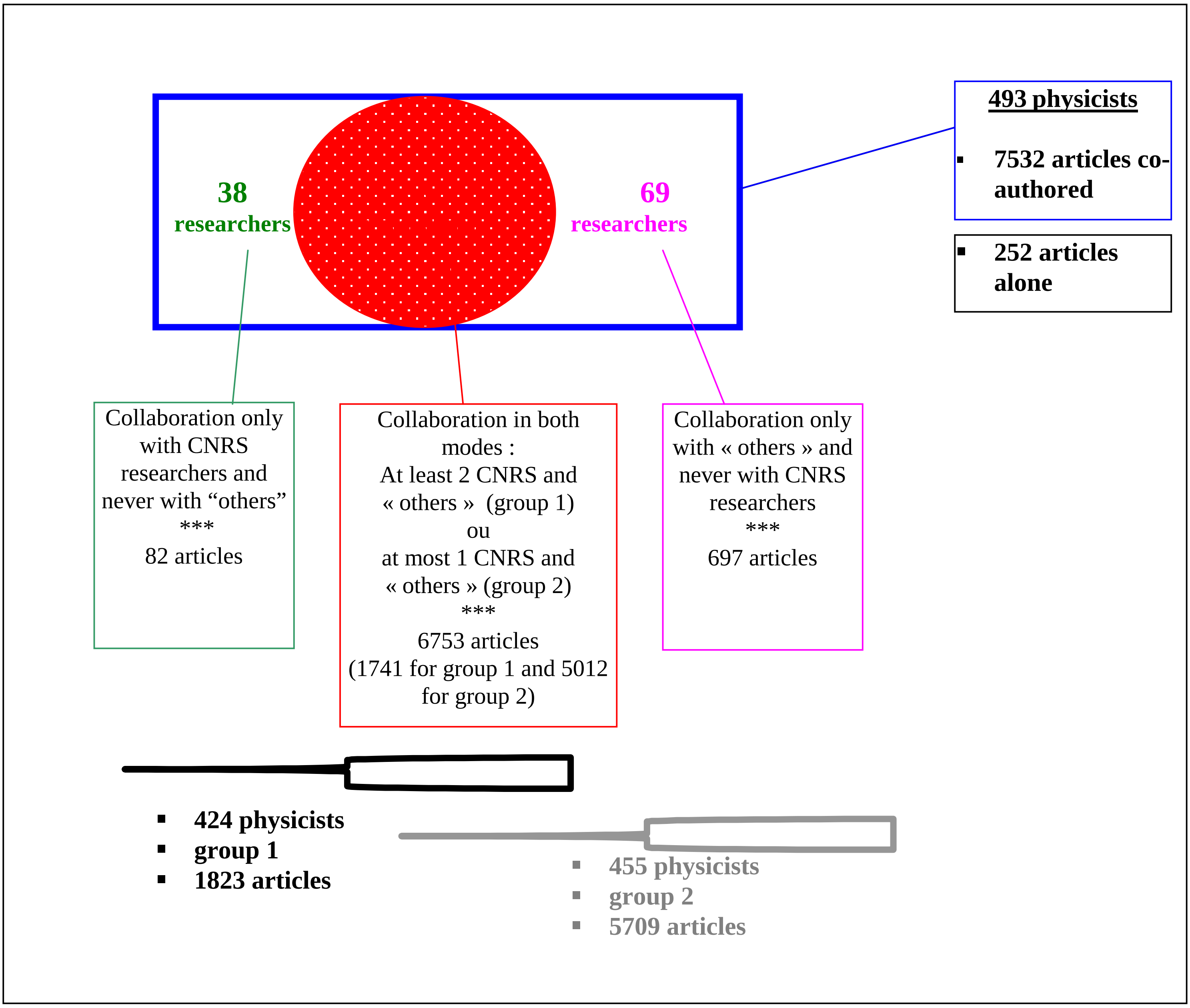
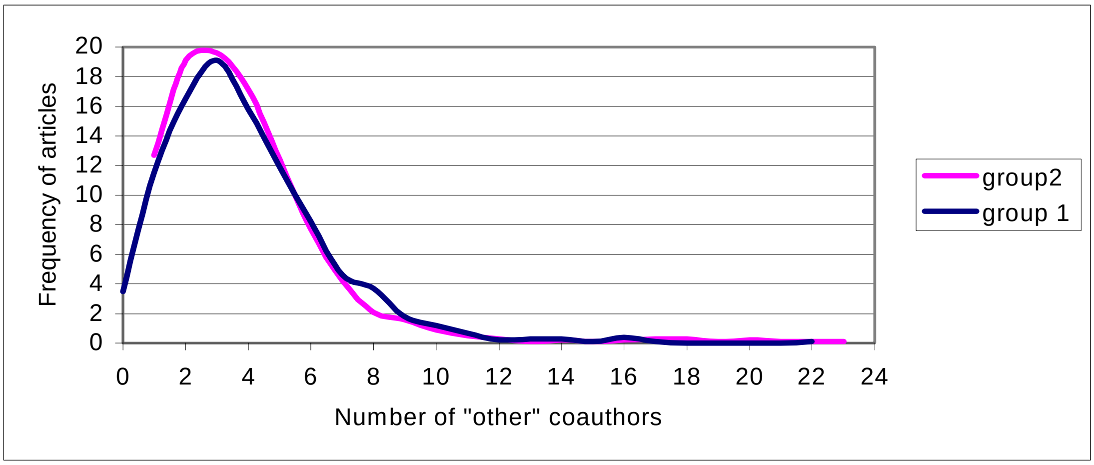

38
researchers
69
researchers
493 physicists
7532 articles co-authored
252 articles alone
Collaboration only with CNRS researchers and never with "others"
***
82 articles
Collaboration in both modes :
At least 2 CNRS and « others » (group 1)
ou
at most 1 CNRS and « others » (group 2)
***
6753 articles
(1741 for group 1 and 5012 for group 2)
Collaboration only with « others » and never with CNRS researchers
***
697 articles
424 physicists
group 1
1823 articles
455 physicists
group 2
5709 articles

Table 1 *Number of co-authored articles published by groups 1 and 2 in relation to the number and type of authors*

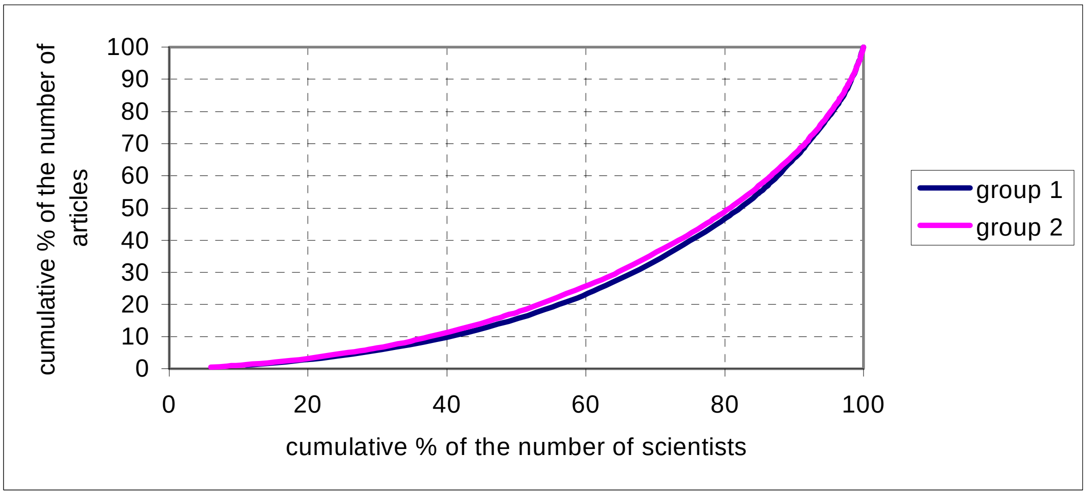
(in brackets: the number of studied CNRS scientists concerned)

Number of co-authored articles	0 « other » scientists	1 « other » scientists	2 « other » scientists	3 « other » scientists	4 « other » scientists	5 « other » scientists	6 « other » scientists	7 « other » scientists	8 « other » scientists	9 « other » scientists	**6.1.1.1 Total**
Group 1 *Including…*	**82 (38)**	**230**	**324**	**375**	**260**	**209**	**127**	**81**	**56**	**161**	**1823 (424***)**
…2 CNRS	**64**	**196**	**268**	**300**	**218**	**172**	**106**	**61**	**47**	**66**	**1498**
…3 CNRS	**15**	**31**	**45**	**60**	**31**	**33**	**15**	**15**	**6**	**7**	**257**
…4 CNRS	**3**	**2**	**9**	**12**	**8**	**4**	**6**	**4**	**3**	**4**	**55**
…5 CNRS et plus	**0**	**1**	**2**	**3**	**3**	**0**	**0**	**1**	**0**	**3**	**13**
Group 2 1 CNRS	*****	**726**	**1087**	**1114**	**976**	**708**	**441**	**241**	**128**	**288**	**5709 (455**)**
Total	82	956	1411	1489	1236	917	568	322	184	367	**7532 (493)**

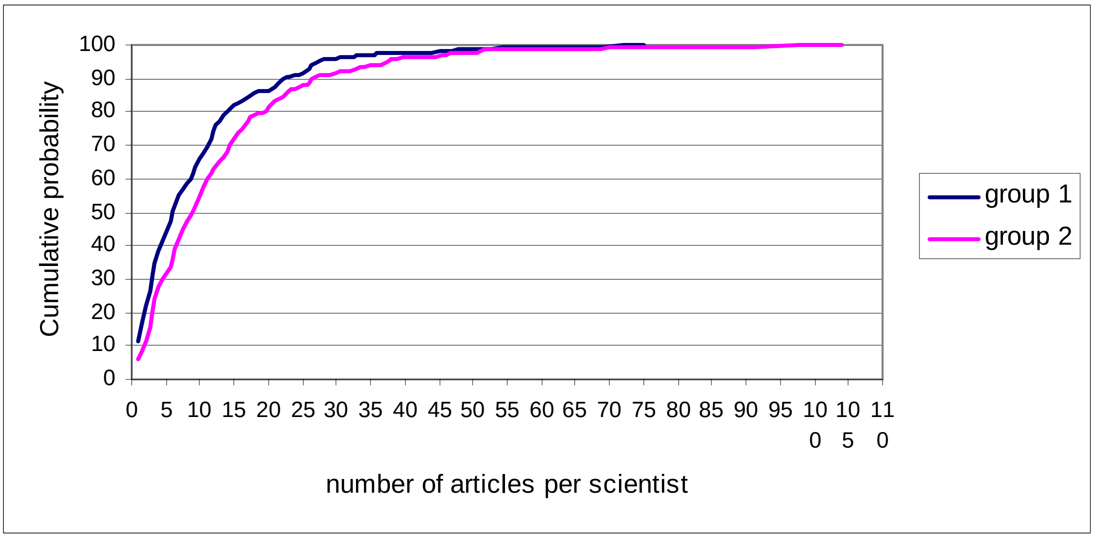
* The number of articles with a single author is 252. They are published by 132 scientists who also have a collaboration activity.
** Including 69 researchers who never published with another CNRS scientists and who contribute for 697 publications to group 1 stock of co-authored papers.
*** Including 38 researchers who never published with « others » and who contribute for 82 publications to group 2 stock of co-authored papers.



Graph 1 *Frequency of the number of articles in relation to the number of "other" co-authors*



Graph 2 *The degree of concentration of the number of articles.*



Graph 3 *Distribution of productivity according to the two forms of collaboration.*

Table 2 *Mean Intra and Inter-towns Intensity of Collaboration*

	Number of scientists	Number of partners towns (*)	Number of articles « intra »	Number of articles « inter » (*)	Intensity intra-town	Mean Intensity between partner towns	Mean Intensity inter-towns
Bagneux	9	6	51	171	62,96	3,45	1,30
Poitiers	11	0	31	0	25,05	0,00	0,00
Gif sur Yvette	16	3	11	40	4,07	0,87	0,16
Grenoble	105	12	666	449	5,42	0,89	0,67
Marseille	18	2	34	18	9,88	0,82	0,10
Meudon	9	2	27	19	33,33	0,52	0,06
Montpellier	20	7	47	83	10,99	0,76	0,33
Orléans	10	0	7	0	6,91	0,00	0,00
Orsay	66	9	174	192	3,61	0,43	0,24
Palaiseau	18	4	15	45	4,36	0,87	0,22
Paris	86	7	249	148	3,03	0,59	0,26
Saint Martin d'Hères	31	5	161	193	15,39	0,93	0,29
Strasbourg	14	2	72	20	35,16	0,89	0,11
Talence	9	0	8	0	9,88	0,00	0,00
Toulouse	29	4	88	63	9,63	1,62	0,41
Villeneuve d'Ascq	10	3	39	31	38,52	3,02	0,57
Villeurbanne	9	2	35	58	43,21	1,43	0,18
Total	**470**	**4**[c]	**1715**[a]	**765**[b]	**18.9** [c]	**1.01** [c]	**0.3** [c]
p	-	-	-	-		*0.0225015*	

* we exclude those couples of towns linked by fewer than five articles.
[a] Each article is weighted by the number of couples that contribute to its publication, otherwise the number of articles would be 1222.
[b] Each article is weighted by the number of couples that contribute to its publication, otherwise the number of articles would be 412.
[c] mean.

Table 3 *The factors studied*

	Number of laboratories per town	Number of scientists	Theoretical number of couple « inter »	Theoretical number of couple « intra »	Publications stock between 1992 and 1997	Mean geographic Distance to partners	Mean Distance to partners in terms of Specialization	Mean Produc-tivity	Mean Quality of the publications	Proportion of articles coauthored by foreigners
Bagneux	1	9	4149	36	328	344	3,02	36,44	3,68	0,12
Poitiers	1	11	5049	55	88	0	5,17	8,00	2,34	0,26
Gif sur Yvette	1	16	7264	120	246	171	3,93	15,38	3,07	0,14
Grenoble	6	105	38325	5460	1870	421	10,19	17,81	3,39	0,50
Marseille	1	18	8136	153	235	361	11,17	13,06	2,84	0,44
Meudon	1	9	4149	36	99	208	5,97	11,00	2,63	0,20
Montpellier	3	20	9000	190	365	548	5,63	18,25	3,47	0,21
Orléans	1	10	4600	45	63	0	8,24	6,30	3,54	0,32
Orsay	3	66	26664	2145	922	334	4,97	13,97	3,69	0,25
Palaiseau	2	18	8136	153	274	297	4,32	15,22	4,77	0,33
Paris	6	86	33024	3655	985	291	6,72	11,45	3,75	0,48
Saint Martin d'Hères	2	31	13609	465	438	418	4,44	14,13	3,78	0,27
Strasbourg	1	14	6384	91	248	449	6,45	17,71	3,69	0,16
Talence	1	9	4149	36	193	0	11,07	21,44	3,94	0,43
Toulouse	2	29	12789	406	379	410	5,30	13,07	2,71	0,24
Villeneuve d'Ascq	1	10	4600	45	184	305	3,12	18,40	4,13	0,28
Villeurbanne	1	9	4149	36	139	188	4,55	15,44	3,02	0,23
Total	34	470	194176	13127	7056	279,12*	6,13*	15,71*	3,44*	0,29*

*mean

Table 4 *Results of the correlations at the town level*

Correlation between	**Intensity intra town, i_I**	**(0 ou 1)[21] (N=136)**	**$i_{IJ} \neq 0$ [22] (N=34)**
Distance	-	-0.09	-0.16
Specialization			
General Profile		-0.017	-0.21
	-		
Physics-Chemistry		-0.18**	-0.14
General Physics		-0.20***	-0.20
Solid-state Physics		-0.14	-0.24
Applied Physics		0.06	-0.08
Materials Science		-0.02	-0.06
Crystallography		0.16*	-0.17
Other		-0.00	-0.16
Size of the scientific community			
▪ *number of researchers*			
N_I	- 0.48 *	-	-
Maximum (N_I, N_J)	-	0.52***	- 0.41**
	-		
	-		
Minimum (N_I, N_J)		0.42***	- 0.31*
Half sum ($N_I + N_J$) /2		0.56***	- 0.45***
▪ *number of possible couples of researchers in the towns under consideration* C_I=N_I * (N_J -1) /2	- 0.39	0.49***	- 0.32*
▪ *stock of publications between 1992 and 1997*			
S_I	- 0.29	-	-
Maximum (S_I,S_J)	-	0.52***	-0.29*
	-		
	-		
Minimum (S_I,S_J)		0.54***	-0.21
Half sum (S_I + S_J)/2		0.60***	-0.33*
Productivity between 1992 and 1997			
P_I	0.62***	-	-
Maximum (P_I,P_J)	-	0.11	0.67***
	-		
	-		
Minimum (P_I,P_J)		0.26***	0.47**
Half sum (P_I + P_J)/2		0.19***	0.69***
Quality of publications			
Q_I	- 0.11	-	-

[21] Existence of a collaboration link inter-towns

[22] Intensity of collaboration between partner towns I and J

Maximum (Q_I,Q_J)	- - -	0.03	0.091
Minimum (Q_I,Q_J)		0.23***	0.033
Half sum (Q_I + Q_J)/2		0.16*	0.075
International Openness			
pe_I	-0.37	-	-
Maximum (pe_I, pe_J)	-	0.14*	-0.26
Minimum (pe_I, pe_J)	-	0.34***	-0.12
Half sum (pe_I + pe_J)/2	-	0.26***	-0.22

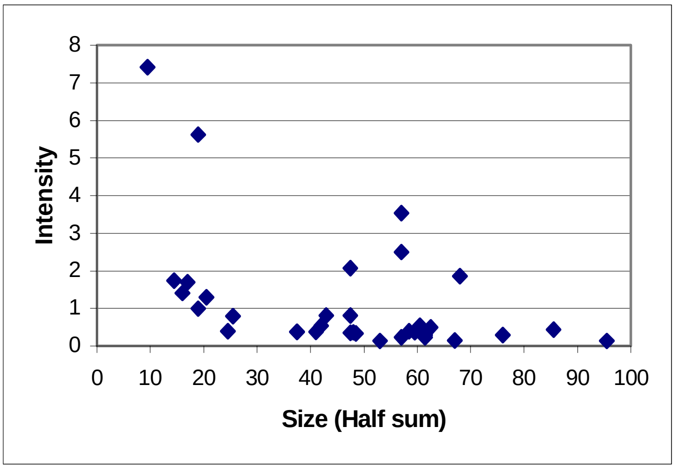


Figure 1 *Intensity of collaboration inter partner towns plotted against the half sum of their size.*

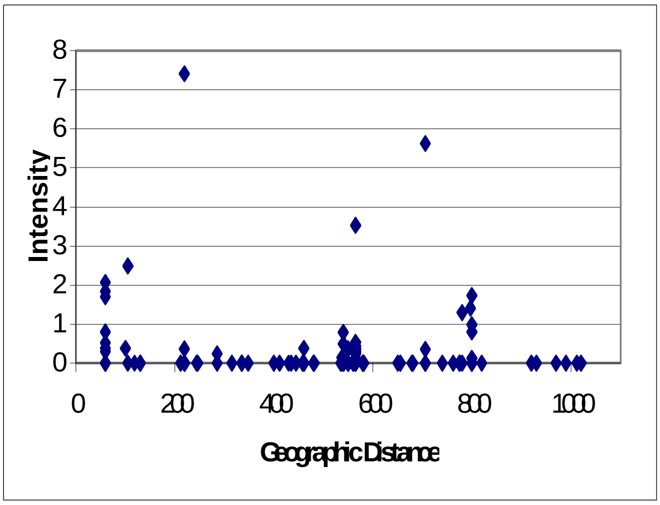


Figure 2 *Intensity of collaboration inter partner towns plotted against the geographic distance.*

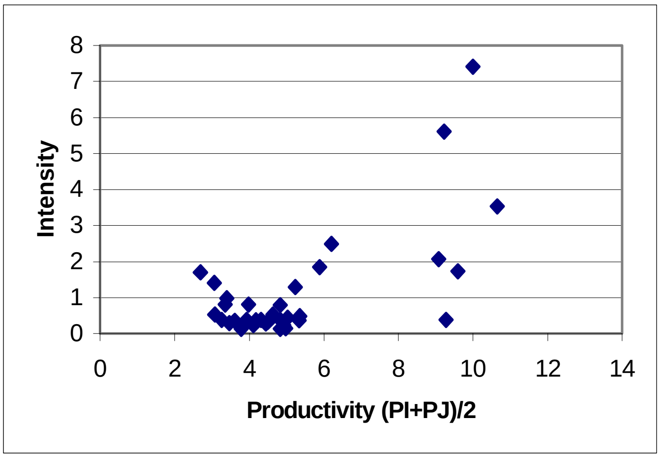


Figure 3 ***Intensity of collaboration inter partner towns plotted against their productivity between 1992 and 1997.***

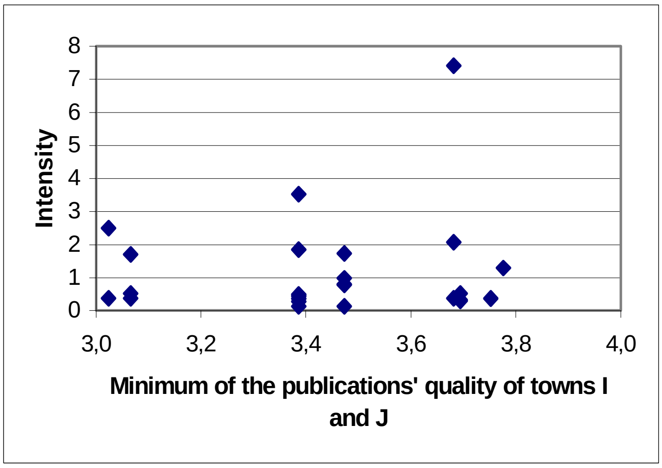


Figure 4 ***Intensity of collaboration inter partner towns plotted against the minimum of their publications' quality.***

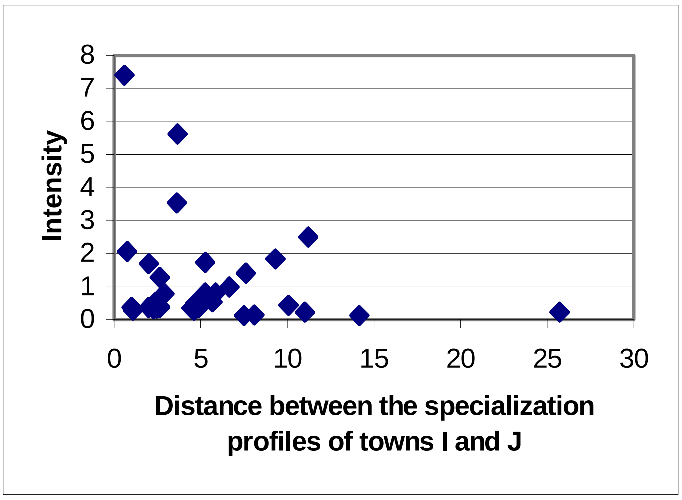


Figure 5 ***Intensity of collaboration inter partner towns plotted against the distance between their specialization profiles.***

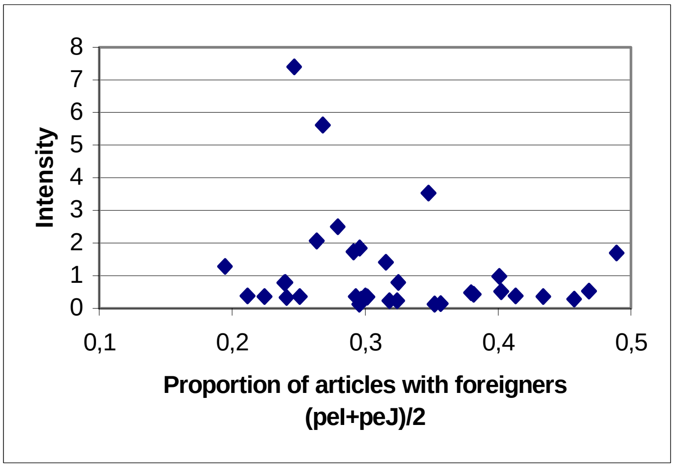


Figure 6 ***Intensity of collaboration inter partner towns plotted against the half sum of the proportion of their articles co-authored by foreign scientists.***

Table 5a *Results of the correlations at the laboratory level*

Correlation between	**Existence of a collaborative link inter-laboratories**	
	Inter towns (N=1044)	**Intra town** (N=78)
Distance	-0.09**	-
Specialization		
General Profile	-0.02	0.40***
Physics-Chemistry	-0.05	0.30***
General Physics	-0.09***	0.35***
Solid-state Physics	-0.04	0.24**
Applied Physics	-0.01	0.05
Materials Science	0.05*	0.41***
Crystallography	0.06**	0.10
Other	-0.03	-0.11
Size of the scientific community		
▪ *Number of researchers*		
N_I	-	-
Maximum (N_I, N_J)	0.21***	0.45***
Minimum (N_I, N_J)	0.22***	0.57***
Half sum ($N_I + N_J$) /2	0.24***	0.57***
▪ *Number of possible couples of researchers in the towns under consideration* $C_I = N_I * (N_J - 1)/2$	0.27***	0.61***
▪ *Stock of publications between 1992 and 1997*		
S_I	-	-
Maximum (S_I,S_J)	0.23***	0.58***
Minimum (S_I,S_J)	0.35***	0.63***
Half sum ($S_I + S_J$)/2	0.31***	0.63***
Productivity between 1992 and 1997		
P_I	-	-
Maximum (P_I,P_J)	0.19***	0.35***
Minimum (P_I,P_J)	0.30***	0.40***
Half sum ($P_I + P_J$)/2	0.26***	0.39***
Quality of publications		
Q_I	-	-
Maximum (Q_I,Q_J)	-0.03	-0.46***
Minimum (Q_I,Q_J)	0.06**	-0.30***
Half sum ($Q_I + Q_J$)/2	0.02	-0.44***
International Openness		
pe_I	-	-
Maximum (pe_I, pe_J)	0.02	0.07

Minimum (pe_I, pe_J)	0.03	-0.06
Half sum (pe_I + pe_J)/2	0.03	0.01

Table 5b *Results of the correlations at the laboratory level*

Correlation between	**Intensity intra labs, i_I** (N=34)	**Intensity of collaboration between partner labs**	
		Inter towns *(N=82)*	Intra towns (N=30)
Distance	-	-0.06	-
Specialization			
General Profile	-	-0.32***	-0.42**
Physics-Chemistry		-0.18	-0.10
General Physics		-0.29***	-0.30
Solid-state Physics		-0.25**	-0.05
Applied Physics		-0.05	-0.26
Materials Science		-0.21*	-0.35*
Crystallography		-0.26**	-0.16
Other		-0.11	-0.26
Size of the scientific community			
▪ *Number of researchers*			
N_I	- 0.42**	-	-
Maximum (N_I, N_J)	-	-0.52***	-0.34*
Minimum (N_I, N_J)	-	-0.51***	-0.02
Half sum ($N_I + N_J$) /2	-	-0.60***	-0.27
▪ *Number of possible couples of researchers in the towns under consideration* $C_I = N_I * (N_J - 1) / 2$	- 0.34**	-0.49***	-0.23
▪ *Stock of publications between 1992 and 1997*			
S_I	- 0.06	-	-
Maximum (S_I,S_J)	-	-0.08	0.18
Minimum (S_I,S_J)	-	-0.14	0.13
Half sum ($S_I + S_J$)/2	-	-0.12	0.17
Productivity between 1992 and 1997			
P_I	0.36**	-	-
Maximum (P_I,P_J)	-	0.51***	0.33*
Minimum (P_I,P_J)	-	0.42***	0.36**

Demi Somme $(P_I + P_J)/2$		0.54***	0.36**
Quality of publications			
Q_I	0.20	-	-
Maximum (Q_I,Q_J)	- - -	-0.03	-0.25
Minimum (Q_I,Q_J)		0.12	0.11
Half sum $(Q_I + Q_J)/2$		0.05	-0.006
International Openness			
pe_I	0.14	-	-
Maximum (pe_I, pe_J)	-	0.34***	-0.005
Minimum (pe_I, pe_J)	-	0.17	0.35*
Half sum $(pe_I + pe_J)/2$	-	0.29***	0.22

Table 6 *Intensity of collaboration and geographic Distance*

Town	Number of laboratories per town	Intensity of collaboration intra laboratory	Mean Intensity of collaboration inter laboratories but intra town	Mean Intensity of collaboration inter laboratories and inter towns
Bagneux	1	58,35	-	2,02
Futuroscope	1	23,22	-	0,00
Gif sur Yvette	1	9,15	-	0,43
Grenoble	6	19,36	2,66	0,52
Marseille	1	11,62	-	0,09
Meudon	1	30,89	-	0,07
Montpellier	3	28,11	0	0,26
Orléans	1	6,41	-	0,00
Orsay	3	33,41	1,38	0,16
Palaiseau	2	11,90	0	0,23
Paris	6	34,53	0,19	0,15
Saint Martin d'Hères	2	37,06	0	0,41
Strasbourg	1	38,02	-	0,11
Talence	1	15,69	-	0,00
Toulouse	2	25,74	1,45	0,21
Villeneuve d'Ascq	1	98,86	-	0,49
Villeurbanne	1	40,05	-	0,38
Total/Mean	34	30,73	0,81	0,32

Annexe 2
The Intensity in the networks